Congratulations! You passed!

Grade received 100% | Latest Submission Grade 100% | To pass 80% or higher

Go to next item

1. A Transformer Network, like its predecessors RNNs, GRUs and LSTMs, can process information one word at a time. (Sequential architecture). 1 / 1 point

- True
- False

Expand

Correct
Correct! A Transformer Network can ingest entire sentences all at the same time.

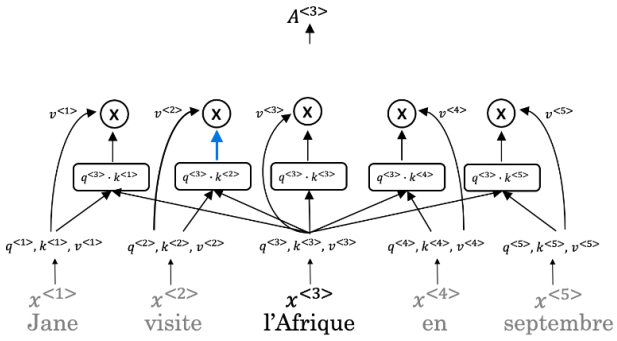
2. Transformer Network methodology is taken from: 1 / 1 point

- Attention Mechanism and RNN style of processing.
- RNN and LSTMs
- Attention Mechanism and CNN style of processing.
- GRUs and LSTMs

Expand

Correct
Transformer architecture combines the use of attention based representations and a CNN convolutional neural network style of processing.

3. What are the key inputs to computing the attention value for each word? 1 / 1 point



- The key inputs to computing the attention value for each word are called the quotation, knowledge, and value.
- The key inputs to computing the attention value for each word are called the query, knowledge, and vector.
- The key inputs to computing the attention value for each word are called the quotation, key, and vector.
- The key inputs to computing the attention value for each word are called the query, key, and value.

Expand

Correct
The key inputs to computing the attention value for each word are called the query, key, and value.

4. What letter does the "?" represent in the following representation of *Attention*? 1 / 1 point

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_?}})V$$

- t
- v
- k
- q

Expand

Correct
k is represented by the $?$ in the representation.

5. Are the following statements true regarding Query (Q), Key (K) and Value (V)? 1 / 1 point

Q = interesting questions about the words in a sentence

K = qualities of words given a Q

V = specific representations of words given a Q

- True
- False

Expand

Correct
Q = interesting questions about the words in a sentence, K = qualities of words given a Q, V = specific representations of words given a Q

6. $\boldsymbol{Attention(W_i^Q Q, W_i^K K, W_i^V V)}$ 1 / 1 point

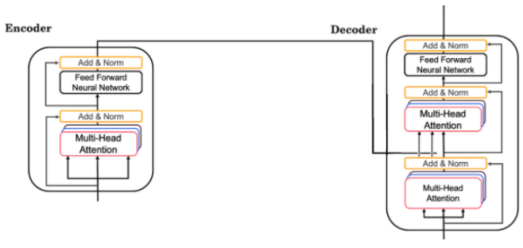
i here represents the computed attention weight matrix associated with the ith "head" (sequence).

- False
- True

Expand

Correct
i here represents the computed attention weight matrix associated with the "head" (sequence).

7. Following is the architecture within a Transformer Network ***(without displaying positional encoding and output layers(s)).*** 1 / 1 point



What is generated from the output of the *Decoder's* first block of *Multi-Head Attention*?

- K

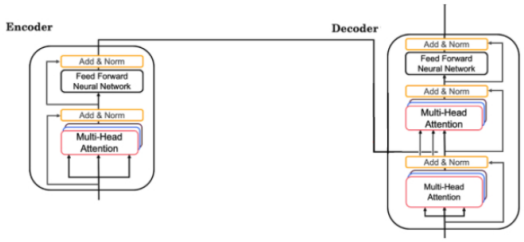
V

Q

Expand

Correct
This first block's output is used to generate the Q matrix for the next Multi-Head Attention block.

8. Following is the architecture within a Transformer Network ***(without displaying positional encoding and output layers(s)).*** 1 / 1 point



What does the output of the *encoder* block contain?

Softmax layer followed by a linear layer.

Contextual semantic embedding and positional encoding information

Linear layer followed by a softmax layer.

Prediction of the next word.

Expand

Correct
The output of the block contains contextual semantic embedding and positional encoding information.

9. Why is positional encoding important in the translation process? (Check all that apply) 1 / 1 point

Position and word order are essential in sentence construction of any language.

Correct

It helps to locate every word within a sentence.

It is used in CNN and works well there.

Providing extra information to our model.

Correct

Expand

Correct
Great, you got all the right answers.

10. Which of these is ***not*** a good criterion for a good positional encoding algorithm? 1 / 1 point

It must be deterministic.

It should output a common encoding for each time-step (word's position in a sentence).

Distance between any two time-steps should be consistent for all sentence lengths.

The algorithm should be able to generalize to longer sentences.

Expand

Correct
This is not a good criterion for a good positional encoding algorithm.